SPEECH RECOGNITION SYSTEM

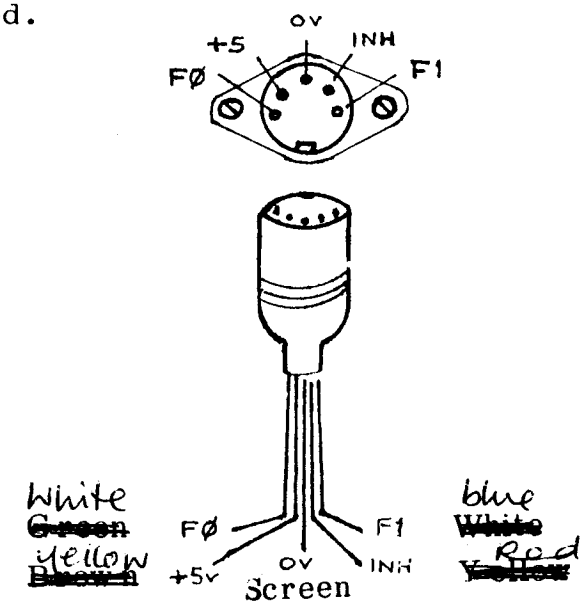
USER MANUAL

Contents

BIG EARS VOICE INTERFACE

1(a) Connection Details for UK101 and OH10 SUPERBOARD

Introduction: Big Ears connects directly to the computer board via its standard 5 PIN DIN socket and the cable supplied.

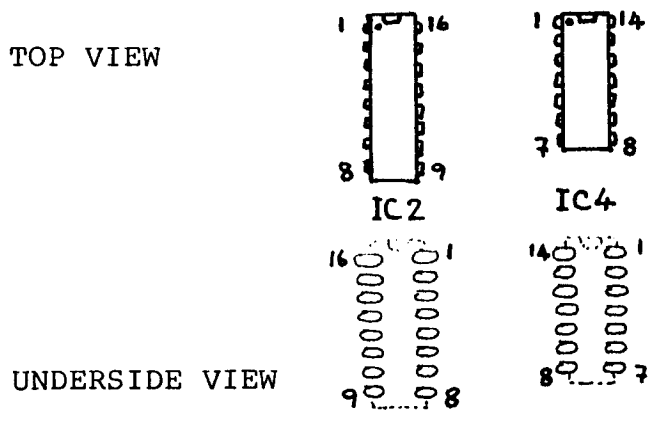


Connections are made as follows:

FUNCTION	DIN CABLE	COMPUTER BOARD (SOLDER TO UNDERSIDE)
+5V	--	IC4 PIN 14 (or any +5V RAIL)
INH	--	IC2 PIN 15
OV	SCREEN	any OV RAIL (or IC4 PIN 7)
FO	--	IC4 PIN 2
F1	--	IC4 PIN 12

Note: please check connections very carefully!

IC pins are numbered thus



2(a) Software for UK101/Superboard

This is supplied as a Basic Program Listing and should be entered and SAVED in two stages:

First enter and SAVE lines 1-19 (this is the part which sets up a short machine code routine) - "Machine Code Loader".

Secondly (after typing NEW) enter and SAVE the remainder. Check carefully for errors - "Analysis Program".

Note that the Machine Code will be loaded into the top of user memory, i.e. the 8th K. It is therefore essential to have memory chips inserted into the last pair of sockets, the minimum amount of memory being 5K in total, inserted as 1, 2, 3, 4 and 8.

Loading Procedure (UK101/Superboard)

1. RESET system

2. MEMORY SIZE ? 7679 [only if full 8K present - if 5K then type 4095]
   (carriage return to all other questions)

3. LOAD
   (load in Machine Code Loader)

4. RUN
   (if the word ERROR is printed, then a mistake
   OK exists in the load Data - check carefully)

5. NEW
   (deletes the Loader)

6. LOAD
   (load in the Analysis Program)

7. RUN

Program Types

LEARN OR TEST?

2(b)

ADDR.	CODE	LABEL	INSTRUCTION	MACHINE CODE FOR Z80 (NASCOM Etc) SYSTEMS
				* CODE & BUFFERS IN NASCOM OCCUPY FROM D00 TO DFF
D00*	3E 4F	START	LD A, ~~4F~~ CF	
	D3 07		OUT (7), A	set Port B to ~~Input~~ control Mode
*	DD 21 80 0D		LD IX, BUFF0	} initialise buffer pointers *
*	FD 21 C0 0D		LD IY, BUFF1	} (values given for 8K operation) *
	16 40		LD D, 64	64 samples
	1E 00		LD E, 00	Clear Flag (word not started)
..10	0E A0	NEXTP	LD C, 160	160 samples / buffer location
..12	DB 05	READ	IN A, (5)	Read Port B
	67		LD H, A	Save
	AD		XOR L	Compare with old bits
	17		RLA	
	30 03		JR NC, FITEST	
	DD 34 00		INC (IX+0)	Count if F0 changed
..1C	17	FITEST	RLA	
	30 03		JR NC, TSTEND	
..1F	FD 34 00		INC (IY+0)	Count if F1 changed
..22	06 0A [or 19]	TSTEND	LD B, COUNT	NB use 0A for 2MHz CPU, use 19 for 4MHz
..24	10 FE	DELAY	DJNZ DELAY	
	6C		LD L, H	Save New as Old
	0D		DEC C	All 160 done?
..28	20 E8		JR NZ READ	No
	CB 03		RLC E	Yes
..2C	20 16		JR NZ NOT1ST	If not start of word
	DD 7E 00		LD A, (IX+0)	
	FD 86 00		ADD (IY+0)	
	37 3F		SCF CCF	Clear carry
	1F		RRA	
	FE 08		CP 8	Threshold
..39	30 08		JR NC, WORDST	Started
	DD 73 00		LD (IX+0), E	
	FD 73 00		LD (IY+0), E	Clear buffers
..41	18 09		JR ROLL	(change to 18 CD to remove screen roll)
..43	1D	WORDST	DEC E	Set Flag
..44	DD 23	NOT1ST	INC IX	
	FD 23		INC IY	
	15		DEC D	All 64 done?
..49	20 C5		JR NZ NEXTP	
	C9		RET	Yes
..4C	3A 60 08	ROLL	LD A, (SCREEN)	
	3C		INC A	
	32 60 08		LD (SCREEN), A	
..53	18 BB		JR NEXTP	

BIG EARS

NASCOM CONNECTIONS
(uses bits 6 & 7 of input port B)

SIGNAL/LINE	NASCOM 2 SIGNAL/LINE	NASCOM 1 SIGNAL/LINE
F0 (~~clear~~ white)	PL4 PIN 5	SKB PIN 8
F1 (~~yellow~~) Blue	PL4 PIN 3	SKB PIN 7
0V. (screen)	PL4 PIN 16	SKB PIN 9
+5V (~~red~~) yellow	PL4 PIN 20	SKB PIN 16
INH (~~blue~~) Red	DO NOT CONNECT	DO NOT CONNECT

SOFTWARE LOADING PROCEDURE (NASCOM)

1/ ENTER ABOVE CODE, USING MONITOR.
2/ SAVE ON TAPE (D00 – D54)
3/ TYPE IN AND SAVE THE BASIC SOFTWARE (LINE 20 ONWARDS), MAKING THE FOLLOWING CHANGES:

```
22   XX=3328
40   Delete
4002 Delete
4005 Delete
4008 DOKE 4100,3328
4041 Delete
```

4/ LOAD THE MACHINE CODE & THE BASIC PROGRAM, AND TYPE RUN.

SPEECH RECOGNITION SYSTEM

Instructions for NASCOM Computers

3. Instructions for using "Big Ears" Speech Recognition System

(Demonstration Software)

1. Set up microphone on table about 1 foot from speaker's mouth and positioned so that it is possible to speak directly into it without turning away from screen.

2. Load Program as indicated in Section 2 and type RUN.

3. The computer will ask

LEARN OR TEST? L	(L selects "learn" mode)
NEW WORD NUMBER? 1	(type a number between 1 & 6)
TYPE IN WORD? APPLES	(type in word 1)
PLEASE SAY APPLES	
NOW!	(say the word loud and clear. Note: one character in the top row of the screen will "spin" until sound has been detected. This indicates that Big Ears is waiting for you and is a good sign that the background noise is low enough).
OKAY	
31 9 0 4 3 1	
0 2 0 1 2 3	(Array of numbers shows how the word has been stored - this is the "voiceprint").
2 2 0 0 0 0	
2 0 0 0 0 0	
1 2 0 0 0 0	
PLEASE SAY APPLES	
NOW!	(The word must be repeated four times. The voiceprint is printed each time).
etc.	
etc.	
LEARN OR TEST? L	
etc.	(Carry on teaching words 2, 3, 4, 5, 6*).
etc.	

4. After 2 or more words have been "taught", you can try out the recognition software.

```
LEARN OR TEST? T

PLEASE SPEAK

NOW!                                    (say one of the words)

OKAY

35 10 3 0 0 0

 4  1 2 2 0 0

 3  4 - - - -                           (voiceprint printed)

 -  - - - - -

 -  - - - - -

APPLES          256.1

PEARS           265.3                   (correlation table printed)

RASPBERRIES     270.3

  - - -          ---

  - - -          ---

  - - -          ---

YOU SAID RASPBERRIES                    (word with highest correlation
                                         is indicated)
```

5. After you have experimented with the system, you can remove the "voiceprint" printout by deleting line 1175.

   The correlation table can be similarly suppressed by deleting lines 2085, 2086 and 2087.

6. If more than 5K of memory is available, the number of words in the vocabulary can be increased. Line 21 sets

   VL = Vocabulary Length, and

   LR = Number of Repetitions when learning.

   (For optimum recognition, set LR = 8 and limit VL to around 10. Extension to a much larger vocabulary is discussed in the Theory Section).

7. Remember that recognition depends on

clarity of speech

similarity of words - very similar words will always be difficult to distinguish. In this respect the vowel content is the dominant feature, thus "pine" and "fine" might be difficult to separate.

8. Line 2090 prints the result of the recognition process . . . "YOU SAID . . ."

Some entertaining effects can be had by changing this to remove the words YOU SAID then, when teaching new words, to type in not the spoken word but the desired "reply". Thus, type in the phrase "I'M A COMPUTER" but repeat (teach) the phrase "WHO ARE YOU". Remember that any word or phrase spoken must last for a maximum of 1 second or it will be incorrectly learned.

UK101 - EXTENDED MONITOR

Warning: When using BIG EARS with the UK101 New Monitor, the following changes are required:

Hardware Connections

F0 to IC5 pin 9
F1 to IC5 pin 5
INH to IC2 pin 16
+5 to IC4 pin 14 (or 5v rail)
0v to IC4 pin 7 (or 0v rail)

Software

```
4    DATA 173,0,223,133,34,69,35,42
5    DATA 144,3,254,128,30,42,144,3
6    DATA 254,192,30,42,144,3,234,234
13   DATA 232,224,64,208,175,96,10668,0
4002 POKE 57088,254
```

4. Theory of Operation

Operation is based on frequency analysis of the first and second formants of the speech waveform. The Interface unit separates the formants and delivers digital pulses to the computer which counts the changes of state (of each formant in each of 64 16 mS sampling periods). This is performed by machine code.

For each period, the two formant counts are then compared against threshold data values to determine which of 5 (formant 2) or 6 (formant 1) frequency ranges are present. The two range indices are now used to determine the location in a two-dimensional (5 x 6) array which will be incremented. This is, therefore, a kind of "frequency-space" and the 64 samples must all fit into it as a 2-dimensional histogram.

When "learning" a word, four or more such histograms are averaged, normalised to have a mean value of zero and a uniform standard deviation. The resulting "voiceprint" is then stored for future correlation.

Software Details

The software is written mainly in subroutines for ease of incorporation into your own applications.

Line Number	Function
4000-4060	"Listen" Subroutine - called by GOSUB 4000. This sets up the call to the machine code (USR) subroutine, enables the hardware interface unit (UK101/Superboard only), clears the input buffers and executes the machine code for real-time voice acquisition. The messages "NOW" and "OKAY" are printed before and after acquisition respectively.
1000-1180	"Classify" subroutine - the two input buffers are processed to produce the 30 element histogram P(30). Line 1175 calls an optional printout of the histogram.
2000-2095	"Correlate" subroutine - the input histogram P(30) is multiplied element by element with each stored Voiceprint and summed to produce correlation results CC (VL). The results are then searched to select the highest value and that word is printed. Lines 2085-2087 give an (optional) printout of all the correlation results. The routine returns with BW set to the word number recognised. This can, of course, be used to take action dependent of the application.

Line Number	Function
3000-3098	"Learn" subroutine - invites the user to create or update his vocabulary. Words must be repeated LR times (LR = 4 to 8) in order to give a statistically good Voiceprint. Voiceprints are stored for each of VL (Vocabulary Length) words; in the two-dimensional 30 x VL array VP (VL, 30). The text strings for each word are stored as array VW$ (VL). The routine returns with BW = word number.
5000-5050	"Pattern Print" Subroutine - can be used to print out the P(30) array, which contains the most recently spoken voice pattern.

Extending to Large Vocabularies

The key to the successful implementation of large vocabularies lies in structuring the application so that the expected response is always one of a reasonably small set of words, with the initial set of words consisting of key words which lead to the next group.

Thus, a Travel reservation system might initially ask "Inland, European or Intercontinental?", to which each of the three replies will lead to a list of, say, 8 or 10 possible destinations. If the destination lists need to be extended, then the word "other" could be included in each one and the program organised to call in the subsequent list.

Program implementation is best achieved as follows. Set VL (Vocabulary Length) in line 20 = total number of words to be stored. Define a control array of (say) 10 elements by adding the line 20 DIM CA (10).

Then change the following lines:

```
2010  FOR Q = 1 TO 10: WD = CA (Q)
2040  NEXT Q
2055  BW = CA (1): BC = CC (CA[1])
2060  FOR Q = 2 TO 10: WD = CA (Q)
2080  NEXT Q
2085)
2086) omit
2087)
```

The correlation routine will now attempt to match only those 10 words whose word numbers are held in array CA (10). The master program (lines 110 to 150 in the demonstration software) must now be modified to set up CA (1) to CA (10) with the "expected" word numbers before asking questions.

A useful hint is to leave certain "master" words permanently in the control array - e.g. "RESET" as word 1 could be used to revert to an initial dialogue no matter where the conversation had reached, and "RUBOUT" as word 2 could be used to allow the speaker another attempt if he sees that his word has been incorrectly recognised. As before, the result of GOSUB 2000 (correlate) is always a printout of the word and BW is set to the word number.

5. BASIC SOFTWARE for Speech Recognition

MACHINE CODE LOADER (SPEECH INPUT) c 1980 Wm Stuart Systems Ltd.
FOR UK101/SUPERBOARD

```
1   REM  SPEECH LOADER (C) 1980 WM STUART SYSTEMS
2   XX=7680
3   DATA 162,  0,134, 33,169,160,133, 32
4   DATA 173,  0,223,133, 34, 69, 35,106
5   DATA 144,  3,254,128, 30,106,144,  3
6   DATA 254,192, 30,106,144,  3,234,234
7   DATA 234,160, 16,136,208,253,165, 34
8   DATA 133, 35,198, 32,208,218,165, 33
9   DATA 208, 30,221,128, 30, 24,125,192
10  DATA  30,238, 32,208, 74,201,  8, 16
11  DATA  13,169,  0,157,128, 30,157,192
12  DATA  30,234,234,234,240,182,230, 33
13  DATA 232,224, 64,208,175, 96,10860, 0
14  CS=0
15  FOR N=0 TO 85
16  READ DD: POKE XX+N,DD: CS=CS+DD
17  NEXT
18  READ DD
19  IF CS<>DD THEN PRINT"ERROR"
```

```
20 REM SPEECH RECOGNITION (C) WM. STUART SYSTEMS 1980

21      VL = 6 : LR = 4
22      XX = 7680 : REM USER MEMORY
24      BØ = XX + 128 : B1 = BØ + 64
25      DIM P(30), CC(VL), VP(VL,30), VW$(VL), PN(30)
30      FOR N = 1 TO 5
32      READ RØ(N), R1(N)
35      NEXT
38      DATA 6, 32, 13, 48, 19, 64, 25, 80, 32, 100
40      POKE 530,1 : REM DISABLE CTRL/C UK101/SUPERBD

100     INPUT "LEARN OR TEST"; A$
105     IF A$ = "L" THEN GOTO 200
110     PRINT "PLEASE SPEAK"
120     GOSUB 4000 : REM LISTEN
140     GOSUB 2000 : REM CORRELATE & PRINT
150     GOTO 100
200     GOSUB 3000 : REM GENERATE
210     GOTO 100

1000    REM CLASSIFY INTO FREQ SPACE
1010    FOR EL = 1 TO 30 : P(EL) = Ø : NEXT
1020    FOR K = Ø TO 63
1030    FØ = 1
1050    IF RØ(FØ) > PEEK(BØ + K) THEN 1100
1060    FØ = FØ + 1
1070    IF FØ < 6 THEN 1050
1100    F1 = 1
1120    IF R1(F1) > PEEK (B1 + K) THEN 1150
1130    F1 = F1 + 1
1140    IF F1 < 5 THEN 1120
1150    EL =(F1-1) * 6 + FØ
1160    P(EL) = P(EL) + 1
1170    NEXT K
1175    GOSUB 5000 : REM (OPTIONAL) PRINT OF FREQ SPACE
1180    RETURN
```

```
2000   REM CORRELATE & IDENTIFY
2010   FOR WD = 1 TO VL
2015   CC(WD) = Ø
2020   FOR EL = 1 TO 30
2030   CC(WD) = CC(WD) + P(EL) * VP(WD,EL)
2035   NEXT EL
2040   NEXT WD
2050   REM NOW FIND BEST
2055   BW = 1 : BC = CC(1)
2060   FOR WD = 2 TO VL
2070   IF CC (WD) < BC THEN GOTO 2080
2075   BW = WD :BC = CC(WD)
2080   NEXT WD
2085   FOR WD = 1 TO VL
2086   PRINT VW$(WD), CC(WD)      OPTIONAL: PRINTS ALL WORD SCORES
2087   NEXT
2090   PRINT : PRINT "YOU SAID"; VW$(BW) : PRINT : PRINT
2095   RETURN

3000   REM VOICEPRINT GEN
3005   INPUT "NEW WORD NUMBER"; WD
3010   INPUT "TYPE IN WORD"; VW$(WD)
3020   FOR EL = 1 TO 30
3025   PN(EL) = Ø
3030   NEXT EL
3040   FOR N = 1 TO LR
3045   PRINT "PLEASE SAY"; VW$(WD)
3050   GOSUB 4000 : PRINT "THANK YOU"
3060   FOR EL = 1 TO 30
3065   PN(EL) = PN(EL) + P(EL)
3070   NEXT EL
3075   NEXT N
3080   S = Ø
3082   FOR EL = 1 TO 30
3084   VP(WD,EL) = PN(EL)/LR - 2.133
3086   S = S + VP(WD,EL) ↑ 2
3088   NEXT EL
```

```
3090    S = SQR(S)
3092    FOR EL = 1 TO 30
3094    VP(WD,EL) = 8 * VP(WD,EL)/S
3096    NEXT EL
3098    RETURN

4000    REM LISTEN
4002    POKE 57088,253 : REM AUDIO ON (UK101 & SUPERBD ONLY)
4005    POKE 11,Ø  [UK101/SUPERBOARD USR ADDRESS]
4008    POKE 12,3Ø  DOKE 400, 3328
4010    FOR X = Ø TO 63
4020    POKE BØ + X, Ø
4025    POKE B1 + X, Ø
4030    NEXT X
4035    PRINT "NOW!"
4040    X = USR(X)
4041    POKE 57088,255 : REM AUDIO OFF (UK101 & C)
4045    PRINT "OKAY"
4050    GOSUB 1000 : REM CLASSIFY
4060    RETURN

5000    FOR R = 1 TO 5 : REM PATTERN PRINT
5005    FOR C = 1 TO 6
5010    PRINT P(C + (R-1) * 6);
5020    NEXT C
5030    PRINT
5040    NEXT R
5050    RETURN
```

BIG EARS - - USER NEWS

Software

Try this modification, which has the effect of reducing the weighting given to the first Voiceprint element. This is the 'all low' count, and corresponds to the silent part of the listening period. By reducing its weight, the non-silent parts are correspondingly accentuated, and the system should be less sensitive to the duration of the words.

```
3076 X=PN(1):PN(1)=PN(1)/4
3077 NS=LR*64-X+PN(1) : AV=(NS/LR)/30
3084 VP(WD,EL)=PN(EL)/LR-AV
2005 P(1)=P(1)/4
```

Note: For good results set LR=8, i.e.
20 VL=11 : LR=8 (11 word vocabulary)

Hardware

The sensitivity of Big Ears is adjustable. To change it, remove the cabinet's lid and turn the small preset potentiometer with a small screwdriver: clockwise to increase, anti-clockwise to decrease. Beware of excess sensitivity. If no word is spoken, the software should listen indefinitely, and if triggered by a very short sound the Voiceprint should show virtually all 64 counts in the top left-hand location. If this is not the case then the sensitivity is too great for the background noise level.